CHEST Critical Care

Interpreting Clinical Trial Results

Christopher Kearney, MD; Brooke Barlow, PharmD; Brandon Pang, MD; and Nicholas A. Bosch, MD

Randomized clinical trials (RCTs) are the gold standard to evaluate intervention efficacy and effectiveness. To apply current, evidence-based interventions to daily practice, it is imperative that practicing intensivists be able to interpret the results of individual RCTs in the context of their patients. In this article, we outline an approach to interpreting critical care RCTs from the perspective of the clinician that focuses on answering four questions: (1) Would my patient have been enrolled and represented in the RCT? (2) Is the intervention feasible? (3) Are there threats to the internal validity of the RCT results? (4) Are the RCT results meaningful? Answers to these four questions can be used to assist intensivists in deciding whether to apply RCT evidence to their patients at the bedside and to avoid common pitfalls of RCT interpretation.

CHEST 2024; 2(4):100097

KEY WORDS: critical care; randomized controlled trials

Why Do Practicing Intensivists Need to Be Able to Interpret Clinical Trials?

Randomized clinical trials (RCTs) are the gold standard to evaluate intervention efficacy and effectiveness. In the field of critical care medicine, landmark RCTs demonstrating intervention efficacy and safety have led directly to several of the routine treatments we use every day, including lung protective ventilation and prone positioning in ARDS,[1,2] corticosteroids in severe pneumonia,[3] and norepinephrine in shock.[4] As important, but less well recognized, are RCT results that have identified harmful interventions that were not adopted into clinical practice.[5]

Evidence generated from RCTs is incorporated routinely into systematic reviews, meta-analyses, and guidelines that inform clinical practice. However, reviews and guidelines may not exist for specific clinical situations, and prolonged delays may occur between publication of RCTs and subsequent updates to guidelines. In addition, summary recommendations resulting from combining the results of many RCTs may miss important contextual factors that limit guideline applicability to specific clinical circumstances. Thus, it is imperative that practicing intensivists be able to interpret the results of individual RCTs in the context of their own patients to apply current, evidence-based interventions to their daily practice. In this article, we outline an approach to interpreting critical care RCTs from the perspective of the bedside clinician that focuses on answering a series of questions to help determine if RCT results should be applied to a specific patient (Fig 1).

ABBREVIATIONS: RCT = randomized controlled trial

AFFILIATIONS: From The Pulmonary Center (C. K., B. P., and N. A. B.), Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA; and the Department of Pharmacy (B. B.), Memorial Hermann The Woodlands Medical Center, The Woodlands, TX.

CORRESPONDENCE TO: Nicholas A. Bosch, MD; email: nabosch@bu.edu

Copyright © 2024 The Authors. Published by Elsevier Inc under license from the American College of Chest Physicians. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

DOI: https://doi.org/10.1016/j.chstcc.2024.100097

SHOUT
SHARE
IT

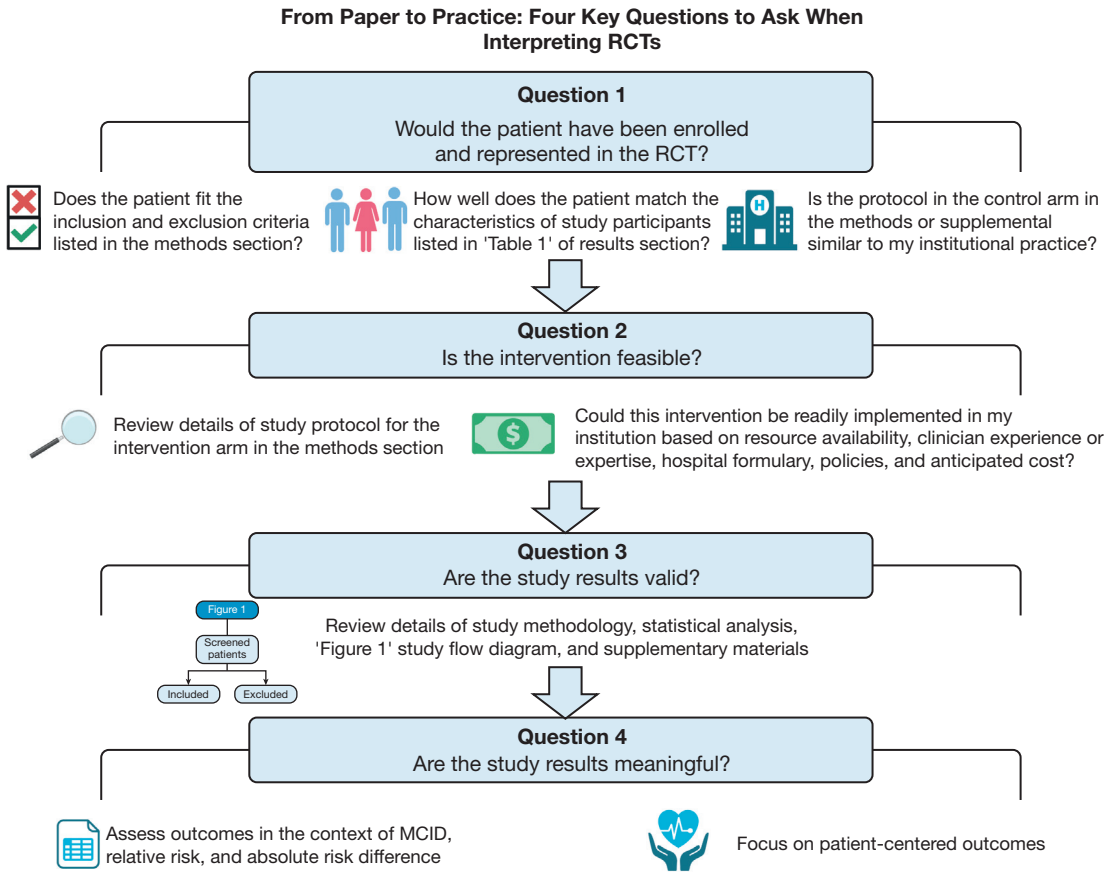
From Paper to Practice: Four Key Questions to Ask When Interpreting RCTs
Question 1
Would the patient have been enrolled and represented in the RCT?
Does the patient fit the inclusion and exclusion criteria listed in the methods section?
How well does the patient match the characteristics of study participants listed in 'Table 1' of results section?
Is the protocol in the control arm in the methods or supplemental similar to my institutional practice?
Question 2
Is the intervention feasible?
Review details of study protocol for the intervention arm in the methods section
Could this intervention be readily implemented in my institution based on resource availability, clinician experience or expertise, hospital formulary, policies, and anticipated cost?
Question 3
Are the study results valid?
Figure 1
Screened patients
Included
Excluded
Review details of study methodology, statistical analysis, 'Figure 1' study flow diagram, and supplementary materials
Question 4
Are the study results meaningful?
Assess outcomes in the context of MCID, relative risk, and absolute risk difference
Focus on patient-centered outcomes

Figure 1 – *Diagram showing questions to ask when interpreting RCTs. MCID = minimal clinically important difference; RCT = randomized clinical trial.*

RCT Terminology

All RCTs prospectively randomize participants to intervention and control groups and then follow participants over time to monitor for the development of outcomes. However, despite these similarities, a plethora of additional methodologic considerations make each RCT unique. These methodologic considerations classify RCTs based on design, treatment assignment strategy, control arm definition, and masking and have important implications for RCT interpretation. Figure 2 includes a glossary of commonly used RCT methodologic terms that can be referenced when reading RCT articles. To enhance the quality and transparency of RCT reporting,[6] these terms are included consistently in RCTs within titles,[7] abstracts,[8] and methods sections.[9]

Question 1: Would the Patient Have Been Enrolled and Represented in the RCT?

Journal clubs—educational sessions focused on interpreting RCT results—often focus on concepts of internal and external validity.[10] RCTs deemed internally valid are those with sound methodologies, procedures, and reporting that, together, minimize bias and increase precision in the results and thus facilitate drawing conclusions about cause and effect. External validity refers to the generalizability of RCT findings to populations and situations that may not have been studied specifically.[11] Although internal validity is a prerequisite to external validity (ie, if RCT findings are unreliable or biased, results will not be externally valid and no reason exists to assess the application of findings to alternative populations), our approach differs and starts with pragmatic questions related to external validity: (1) If the patient had been approached to participate in the RCT, would they have been eligible potentially? and (2) Is the patient represented by the participants enrolled in the RCT?

By screening RCTs using these questions of external validity, clinicians can decide, on first pass, whether the clinical circumstances of the patient match closely enough with the population included in the RCT to decide whether to proceed with additional RCT review. To answer questions related to external validity, we examined the methods section and supplement to

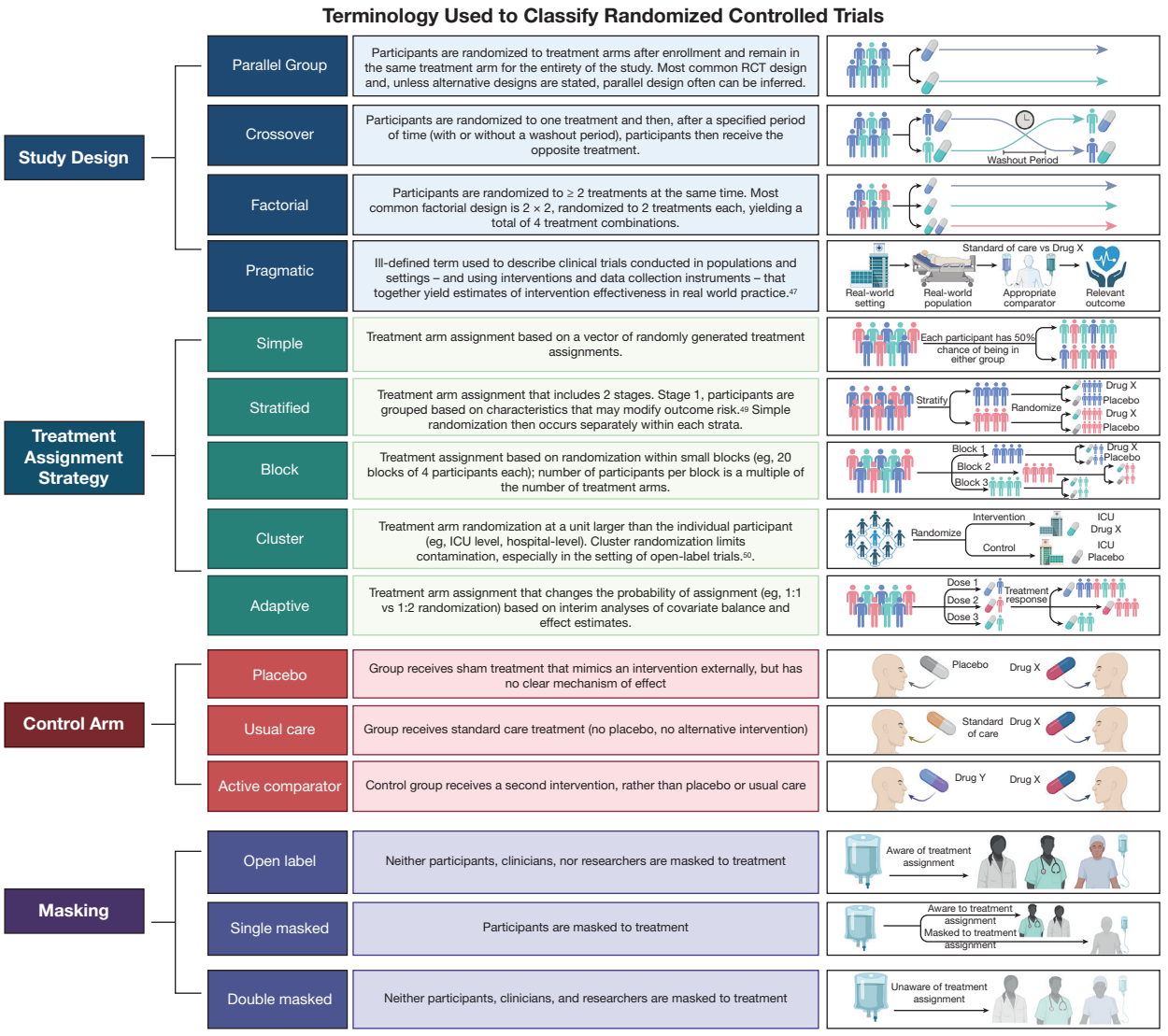
Terminology Used to Classify Randomized Controlled Trials
Study Design
Parallel Group
Participants are randomized to treatment arms after enrollment and remain in the same treatment arm for the entirety of the study. Most common RCT design and, unless alternative designs are stated, parallel design often can be inferred.
Crossover
Participants are randomized to one treatment and then, after a specified period of time (with or without a washout period), participants then receive the opposite treatment.
Washout Period
Factorial
Participants are randomized to ≥ 2 treatments at the same time. Most common factorial design is 2 × 2, randomized to 2 treatments each, yielding a total of 4 treatment combinations.
Pragmatic
Ill-defined term used to describe clinical trials conducted in populations and settings – and using interventions and data collection instruments – that together yield estimates of intervention effectiveness in real world practice.47
Standard of care vs Drug X
Real-world setting
Real-world population
Appropriate comparator
Relevant outcome
Treatment Assignment Strategy
Simple
Treatment arm assignment based on a vector of randomly generated treatment assignments.
Each participant has 50% chance of being in either group
Stratified
Treatment arm assignment that includes 2 stages. Stage 1, participants are grouped based on characteristics that may modify outcome risk.49 Simple randomization then occurs separately within each strata.
Stratify
Randomize
Drug X
Placebo
Drug X
Placebo
Block
Treatment assignment based on randomization within small blocks (eg, 20 blocks of 4 participants each); number of participants per block is a multiple of the number of treatment arms.
Block 1
Block 2
Block 3
Drug X
Placebo
Cluster
Treatment arm randomization at a unit larger than the individual participant (eg, ICU level, hospital-level). Cluster randomization limits contamination, especially in the setting of open-label trials.50.
Randomize
Intervention
Control
ICU
Drug X
ICU
Placebo
Adaptive
Treatment arm assignment that changes the probability of assignment (eg, 1:1 vs 1:2 randomization) based on interim analyses of covariate balance and effect estimates.
Dose 1
Dose 2
Dose 3
Treatment response
Control Arm
Placebo
Group receives sham treatment that mimics an intervention externally, but has no clear mechanism of effect
Placebo
Drug X
Usual care
Group receives standard care treatment (no placebo, no alternative intervention)
Standard of care
Drug X
Active comparator
Control group receives a second intervention, rather than placebo or usual care
Drug Y
Drug X
Masking
Open label
Neither participants, clinicians, nor researchers are masked to treatment
Aware of treatment assignment
Single masked
Participants are masked to treatment
Aware to treatment assignment
Masked to treatment assignment
Double masked
Neither participants, clinicians, and researchers are masked to treatment
Unaware of treatment assignment

Figure 2 – *Diagram showing randomized clinical trial methodological considerations.*

identify eligibility criteria and protocolized treatments that are not specifically part of the intervention or control arm and examined the table of patient characteristics at the beginning of the results section (often referred to as a "Table 1"). The greater the discrepancies between the patient characteristics and RCT enrollment criteria, the more likely we are to conclude that an RCT is not applicable to the patient. In our experience, common discrepancies between enrollment criteria and patient characteristics that limit generalizability include time from critical illness onset and severity of illness. Participants in RCTs often are enrolled at the beginning of a hospitalization.[3,8,12] However, patients who are past the acute phase of critical illness (ie, recovery or prolonged critical illness and mechanical ventilation)[13] may have altered immune responses[14] and neuroendocrine hormone balance[15] that may explain why some interventions with mortality benefit in early critical illness[16] instead may demonstrate harm at later time points.[17] Responses to treatments may be heterogenous based on severity of illness,[18] and thus we generally avoid extrapolating interventions to populations that were not examined in the original RCT contexts (eg, angiotensin II as a "salvage vasopressor"[19]).

Even in settings where the patient would meet enrollment criteria, reviewing patient characteristics in a Table 1[6] allows us to determine if the patient is well represented in the study population. We exercise caution in applying RCT evidence if very few participants are enrolled who have similar characteristics to a patient we are treating, even if that patient would have been eligible for enrollment. For example, although the eligibility criteria in the Activated Protein C and Corticosteroids

TABLE 1] Assessing the Interval Validity of RCTs

RCT Component	Questions We Ask to Assess Internal Validity	Location Within RCT Article
Concealment and masking	• Which of the following groups were masked to the randomization sequence (concealment) and treatment (masking): participants, clinicians, researchers? • If the RCT was open-label, did differential exclusion occur before treatment initiation as part of modified intention-to-treat analyses? Differential loss to follow-up? Differential outcome missingness?	Methods
Fidelity to the intervention	• Did large crossover occur between treatment arms that could decrease ability to detect a difference? • Did participants randomized to a treatment arm mostly receive that treatment? • If participants did not adhere to an intervention, why not? Could low adherence be expected in if the intervention were implemented into clinical practice?	Results; CONSORT study flow diagram[6]
Loss to follow-up	• Was the number of participants who were lost to follow-up reported? • How were participants who did not experience an outcome handled at the last follow-up? • Were any patients lost to follow-up before initiating the assigned treatment? • How did analyses account for loss to follow-up?	Methods: outcomes and statistical analysis plan sections; Results (before outcome reporting)
Measurement ascertainment and missing data	• Were any outcomes a subjective scoring system that could yield biased results if researchers were unmasked (ie, open-label)? • Are outcomes based on routine health collected data or manual extraction? • What percentage of participants had missing outcome data? Were missing outcomes more common in one treatment group than the other? • How were missing outcome data handled in analyses?	Methods: outcome and procedures sections
Statistical analyses	• Were intention-to-treat, modified intention-to-treat, or per-protocol principles used in the primary analyses? • If per-protocol principles were used, was the mechanism of treatment adherence considered in analyses?	Methods: statistical analysis plan sections
Protocol	• Were the statistical analysis plan and primary outcome prespecified? • Did the study have protocol changes after enrollment started? If yes, did changes to the primary outcome or to the planned sample size occur? • Was the study terminated early? If yes, why?	Methods and supplement

CONSORT = Consolidated Standards of Reporting Trials; RCT = randomized controlled trial.

for Human Septic Shock RCT[9] required potential participants to be receiving at least 0.25 μg/kg/min of norepinephrine (or an equivalent dose of another vasopressor), the mean observed dose of norepinephrine at enrollment of included patients was > 1 μg/kg/min, a dose much higher than that typically used at the time of corticosteroid initiation in patients with septic shock in the United States.[20]

Related to the concepts of whether patients we are treating would be enrolled or represented in our study is whether the standard of care being applied to the patients is similar to the standard-of-care treatments applied during the conduct of the RCT. When standard practices in RCTs differ greatly from those we use in our own practices, it often is difficult for us to establish the generalizability of RCT findings to patients we are treating because the underlying care they receive is different. Clues to potentially large differences in standard of care include RCTs that were conducted before major changes to guidelines, when RCT settings (eg, country, type of ICU, resource availability) differ

greatly from our own practice, and when baseline cotreatments listed in Table 1 seem dissimilar to those treatments being used in the patients we are treating. As a final note, when control arm usual care does not represent current standard therapy,[21] large effect estimates observed between intervention and control arms in clinical trials[1] may be decreased in clinical contexts.

Question 2: Is the Intervention Feasible?

The second question of external validity we ask ourselves also is a question of practicality and is used to determine if RCT interventions are feasible in our clinical settings. The feasibility of applying an RCT intervention locally depends on resource availability, costs, and complexity of the intervention. Although adapting RCT interventions to promote real-world feasibility in clinical contexts is a core component of implementation science, the effectiveness of an adaptation may depend on context and the degree to which it differs from the RCT intervention.[22] Applying only one part of a complex intervention or substituting a related, alternative intervention (eg, different medication in the same class) may not result in recapitulation of observed RCT effects. For example, applying shorter sessions of prone positioning than the 16 hours found to reduce mortality in the landmark Prone Positioning in Severe Acute Respiratory Distress Syndrome RCT[2] does not yield similar mortality benefits.[23] Other examples have less clear implications: (1) Does using hydrocortisone alone yield similar effects on mortality to those observed with combination hydrocortisone and fludrocortisone in the Activated Protein C and Corticosteroids for Human Septic Shock study?[9] and (2) Does application of the mantra "dry lungs equal happy lungs" by using diuresis to avoid hypervolemia yield similar effects to the complex conservative management protocol to which patients were randomized in the Fluids and Catheters Treatment Trial RCT[24] that included frequent assessments of central venous and pulmonary artery occlusion pressure, dobutamine, and urine output in addition to diuretics? In more straightforward cases, resource limitations, lack of expertise, or hospital policy may prevent the use of expensive or resource-intensive therapies at all, such as angiotensin II,[19] extracorporeal membrane oxygenation,[25] and prone positioning.[26]

Question 3: Does the Internal Validity of the RCT Results Face Threats?

As soon as we have established if an RCT population and intervention are potentially applicable to and feasible for a patient in our care, we then ask ourselves if threats to the internal validity of the RCT exist. That is, do the design or conduct related to randomization, concealment or masking, loss to follow-up, measurement ascertainment, fidelity to the intervention, or the statistical analysis plan have potential flaws that could lead to biased and unreliable results?[27]

Tools exist to detect and quantify RCT bias risks systematically.[28] A comprehensive assessment of internal validity is beyond the scope of this review and likely is impractical for practicing clinicians. Thus, we instead outlined components of RCTs that we used to evaluate quickly threats to internal validity in Table 1 and discuss below (1) our general approach and (2) nuances to internal validity in critical care contexts.

We begin with a focused search of the methods section to identify design features (Fig 2) that may increase the risk of bias and that may lead us to question internal validity. In our experience, these design features are often related to (1) allocation concealment (hiding the randomization sequence from investigators and clinicians through sealed envelopes or central randomization), (2) masking (hiding the assigned treatment arm from participants and investigators [double-masked] and statistical analysts [triple-masked]), (3) outcomes that include subjective components (eg, scoring systems for dyspnea, functional status), (4) the approach to analyses in the case of incomplete adherence or follow-up (eg, intention to treat), and (5) deviations or changes to the study protocol, especially those made in response to slow or low enrollment or those included only in trial registries, but not in published protocols.[29] In some cases, a single design feature that increases bias risk may not reach our threshold of concern unless coupled with other at-risk features. For example, the risk of ascertainment bias in an open-label (ie, unmasked) study may hinge on whether the primary outcome was a subjective score.[30] After identifying design features that increase risk of systemic bias, we use the results section (typically the first two paragraphs focused on study enrollment and exclusions, loss to follow-up, missing data, and treatment adherence) to further fine-tune assessments of internal validity. For example, risks of bias related to a protocol change to the primary outcome may depend on whether a study was terminated early or whether the results reported the primary outcome stratified by the original and new definition.

In addition to the above overview summary of assessing threats to internal validity, we call to the readers'

attention specific concepts in critical care RCTs that warrant additional discussion, as follows.

Modified Intention-to-Treat Principles

The key benefit to RCTs is randomization to an intervention or control arm that minimizes both measured and—importantly—unmeasured confounding. Differences between traditional intention-to-treat and per-protocol analyses have been described in detail previously.[31] In summary, intention-to-treat analyses preserve the original separation of treatment arms based on randomization, and thus maintain the benefits of randomization itself (minimization of unmeasured confounding), but also may yield smaller differences in observed outcomes between treatment arms. Per-protocol analyses—analyses among participants who adhered to the assigned treatment arm—yield efficacy effects (effects of interventions under ideal circumstances), but results may be biased if the mechanisms of nonadherence are not accounted for explicitly in analyses.

RCTs increasingly are analyzed using a hybrid principle—modified intention to treat—that combines some aspects of intention-to-treat and per-protocol principles.[32] The definition of modified intention to treat varies from trial to trial, but most RCTs using modified intention-to-treat analyses exclude participants after randomization but before treatment initiation. These postrandomization exclusions, like per-protocol analyses, potentially can lead to biased results if pretreatment exclusion criteria systematically differ between treatment groups.[33] To identify studies at risk of potentially biased results from using modified intention-to-treat principles, we review the methods section to confirm that exclusion criteria between randomization and intervention initiation were made a priori and that such exclusions similarly were likely in both treatment and control arms. RCTs where the participant is unmasked or where the control group is usual care may be particularly at risk of bias.[33] We also review Consolidated Standards of Reporting Trials[6] study flow diagrams in the results—commonly Figure 1 in RCT articles—to identify how many participants were excluded in each arm and whether large differences exist in postrandomization, pretreatment initiation exclusions. Although small numbers and balance in exclusions do not guarantee unbiased modified intention-to-treat analyses, we generally feel more confident in the validity of the results when these criteria hold true.

Early Stopping of RCTs

RCTs are stopped early for a variety of reasons, including reaching prespecified harm, futility or benefit thresholds in interim analyses, and slow enrollment.[34] When RCTs are stopped early, by definition they do not reach their prespecified sample size. Although the rationale for stopping early based on benefit or harm has face validity (ie, it may be suspect ethically to randomize patients to the control arm if an intervention is concluded to have efficacy), trials stopped early for benefit yield overly large effect estimates compared with trials that are not stopped early.[35] In addition, when these overestimated effects are published, it may decrease the likelihood that future studies attempt to replicate such results.[36] During the COVID-19 pandemic, many RCTs were stopped early for slow enrollment.[37,38] When enrollment is lower than predicted, funding sources dry up or exogenous factors (ie, the end of the pandemic) lead to enrollment changes, study investigators may need to stop enrollment before reaching the planned sample size. Although guidelines have been developed to address trial modifications in the setting of extenuating circumstances,[39] publication bias of RCTs stopped early (ie, systematic higher likelihood of publishing positive or statistically significant results)[34] may lead to publication of chance differences in intervention effectiveness that, if the RCT would not have been stopped early, would have regressed to the mean. Alternatively, if published, trials stopped early for slow enrollment that did not show differences in outcomes at the time of trial cessation may lead to potentially incorrect interpretations that the tested intervention was not efficacious. Finally, effect estimates from trials that do not reach their estimated required sample size may have low precision (ie, wide CIs) that make interpretation challenging. Thus, we exert caution when applying interventions from RCTs that stopped early because of benefit or slow enrollment, especially when RCTs are small or when no additional RCTs have been performed yet to replicate findings.

The Competing Risk of Death

Patients who are critically ill, by definition, are at high risk of imminent death.[40] This high risk of death creates methodologic challenges—termed *the competing risk of death*[41]—for studies when outcomes of interest are not mortality. For example, an RCT that compares intervention A with intervention B for a primary outcome of hospital length of stay might conclude that intervention A has shorter length of stay compared with

intervention B when, in reality, intervention A actually increased early mortality. Because those who received intervention A died quickly, they experienced shorter lengths of stay compared with those who received intervention B. Many ways to account for the competing risk of death in outcome selection and analysis exist,[42] including using time-to-event modeling approaches (eg, Fine-Grey models)[43] that explicitly account for competing risks and using free-day composite outcomes that combine duration-related outcomes and death into a single score.[44] When we are interpreting critical care RCTs for which the primary outcome is not mortality or not a free-day composite outcome, we review the statistical analysis plan to see if time-to-event models that account for competing risks were used and if secondary outcomes included mortality. We avoid implementing RCT evidence in circumstances where competing risks were not accounted for, the interventions had risk, and either mortality was not reported or mortality effect estimates suggested potential harm.[45]

Question 4: Are the RCT Results Meaningful?

The last question we ask ourselves is one of meaningfulness. Results generally are meaningful when outcomes are patient centered and effect estimates (ie, measurements of differences in outcomes between treatment arms) are clinically significant. Patient-centered outcomes are those of value and importance to patients[46] and often are related to survival, physical function, quality of life, symptom control, and adverse events.[47] Clinically significant effects are those that have impact and are important to clinical practice and may include estimates for patient-centered outcomes such as death and survival, but also for less patient-centered outcomes such as mechanical ventilation duration and ICU-free days.[24] Traditionally, critical care trials have been powered to identify only unrealistically large treatment effects,[48] and thus most so-called positive practice changing critical care trials have yielded large, statistically significant effect estimates. However, statistical significance does not necessarily imply clinical significance, especially in the setting of small, precise effect estimates or when primary outcomes are intermediate process outcomes.[49] In cases where a given effect estimate has unclear clinical significance, minimal clinically important difference thresholds may be useful, but it should be noted that these are both disease and outcome specific.[50] We are most likely to implement RCT results into practice when outcomes are patient centered and effects are clinically significant.

Conclusions

In this article, we propose a series of questions to aid intensivists in the decision of whether to apply RCT evidence to patients at the bedside and to avoid common pitfalls of RCT interpretation. We recognize that answers to these questions are almost never "yes" or "no," but instead are "probably" or "probably not." Thus, decisions about applying RCT evidence into practice—like clinical medicine—remain a personal judgement decision about the external and internal validity of the findings and the feasibility of the intervention itself within local contexts.

Financial/Nonfinancial Disclosures

None declared.

References

1. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med*. 2000;342(18):1301-1308.
2. Guérin C, Reignier J, Richard J-C, et al. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med*. 2013;368(23):2159-2168.
3. Dequin P-F, Meziani F, Quenot J-P, et al. Hydrocortisone in severe community-acquired pneumonia. *N Engl J Med*. 2023;388(21):1931-1941.
4. De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med*. 2010;362(9):779-789.
5. Santacruz CA, Pereira AJ, Celis E, Vincent J-L. Which multicenter randomized controlled trials in critical care medicine have shown reduced mortality? A systematic review. *Crit Care Med*. 2019;47(12):1680.
6. Consolidated Standards of Reporting Trials. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. EQUATOR network website. Accessed May 10, 2024. https://www.equator-network.org/reporting-guidelines/consort/
7. Hébert PC, Wells G, Blajchman MA, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *N Engl J Med*. 1999;340(6):409-417.
8. Papazian L, Forel J-M, Gacouin A, et al. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med*. 2010;363(12):1107-1116.
9. Annane D, Renault A, Brun-Buisson C, et al. Hydrocortisone plus fludrocortisone for adults with septic shock. *N Engl J Med*. 2018;378(9):809-818.
10. Arora S, Schriger DL. Journal club: frequent fliers, internal and external validity, and the problem with making continuous variables binary. *Ann Emerg Med*. 2009;53(5):594-595.
11. Patino CM, Ferreira JC. Internal and external validity: can you apply research study results to your patients? *J Bras Pneumol*. 2018;44(3):183.
12. National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network, Shapiro NI, Douglas IS, et al. Early restrictive or liberal fluid management for sepsis-induced hypotension. *N Engl J Med*. 2023;388(6):499-510.
13. MacIntyre NR, Epstein SK, Carson S, et al. Management of patients requiring prolonged mechanical ventilation: report of a NAMDRC consensus conference. *Chest*. 2005;128(6):3937-3954.

14. Kalb TH, Lorin S. Infection in the chronically critically ill: unique risk profile in a newly defined population. *Crit Care Clin*. 2002;18(3):529-552.
15. Van den Berghe GH. Acute and prolonged critical illness are two distinct neuroendocrine paradigms. *Verh K Acad Geneeskd Belg*. 1998;60(6):487-518. ; discussion 518-520.
16. Villar J, Ferrando C, Martínez D, et al. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir Med*. 2020;8(3):267-276.
17. Steinberg KP, Hudson LD, Goodman RB, et al. Efficacy and safety of corticosteroids for persistent acute respiratory distress syndrome. *N Engl J Med*. 2006;354(16):1671-1684.
18. Munshi L, Del Sorbo L, Adhikari NKJ, et al. Prone position for acute respiratory distress syndrome. A systematic review and meta-analysis. *Ann Am Thorac Soc*. 2017;14(suppl 4):S280-S288.
19. Vail EA, Bosch NA, Law AC, Gershengorn HB, Wunsch H, Walkey AJ. Adoption of a novel vasopressor agent in critically ill adults. *Ann Am Thorac Soc*. 2023;20(11):1662-1667.
20. Bosch NA, Teja B, Wunsch H, Walkey AJ. Practice patterns in the initiation of secondary vasopressors and adjunctive corticosteroids during septic shock in the US. *Ann Am Thorac Soc*. 2021;18(12). https://www.atsjournals.org/doi/abs/10.1513/AnnalsATS.202102-196OC
21. Slutsky AS, Marco Ranieri V. Mechanical ventilation: lessons from the ARDSNet trial. *Respir Res*. 2000;1(2):73-77.
22. Moore G, Campbell M, Copeland L, et al. Adapting interventions to new contexts—the ADAPT guidance. *BMJ*. 2021;374:n1679.
23. Bloomfield R, Noble DW, Sudlow A. Prone position for acute respiratory failure in adults. *Cochrane Database Syst Rev*. 2015;2015(11):CD008095.
24. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, Wiedemann HP, Wheeler AP, Bernard GR, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med*. 2006;354(24):2564-2575.
25. Tillmann BW, Pham T, Scales DC, Fan E, Pinto R, Rubenfeld G. Impact of center of admission on receipt of extracorporeal membrane oxygenation among patients with hypoxemic respiratory failure: a secondary analysis of the Large Observational Study to Understand the Global Impact of Severe Acute Respiratory Failure and Reseau Europeen de Recherche en Ventilation Artificielle Influenza A(H1N1) Studies. *Chest Crit Care*. 2024;2(1):100040.
26. Law AC, Forbath N, O'Donoghue S, Stevens JP, Walkey AJ. Hospital-level availability of prone positioning in Massachusetts ICUs. *Am J Respir Crit Care Med*. 2020;201(8):1006-1008.
27. United States Preventative Serveices Task Force. Procedure manual appendix VI. Criteria for assessing internal validity of individual studies. United States Preventive Services Taskforce website. Accessed May 10, 2024. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-appendix-vi-criteria-assessing-internal-validity-individual-studies
28. Cochrane Collaboration. RoB 2: a revised Cochrane risk-of-bias tool for randomized trials. Cochrane Bias website. Accessed May 10, 2024. https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials
29. Ewart R, Lausen H, Millian N. Undisclosed changes in outcomes in randomized controlled trials: an observational study. *Ann Fam Med*. 2009;7(6):542-546.
30. Catalog of Bias. Ascertainment bias. Catalog of Bias website. 2017. Accessed May 10, 2024. https://catalogofbias.org/biases/ascertainment-bias/
31. Santos-Gallego CG, Requena-Ibanez JA, Badimon J. Per-protocol versus intention-to-treat in clinical trials: the example of GLOBAL-LEADERS trial. *J Am Heart Assoc*. 2022;11(10):e025561.
32. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*. 2010;340:c2697.
33. Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. *Clin Trials Lond Engl*. 2023;20(3):269-275.
34. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA*. 2014;311(10):1045-1052.
35. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303(12):1180-1187.
36. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. *BMJ*. 2012;344:e3863.
37. Audisio K, Lia H, Robinson NB, et al. Impact of the COVID-19 pandemic on non-COVID-19 clinical trials. *J Cardiovasc Dev Dis*. 2022;9(1):19.
38. Gottlieb RL, Vaca Carlos E, Paredes R, et al. Early remdesivir to prevent progression to severe Covid-19 in outpatients. *N Engl J Med*. 2022;386(4):305-315.
39. Orkin AM, Gill PJ, Ghersi D, et al. Guidelines for reporting trial protocols and completed trials modified due to the COVID-19 pandemic and other extenuating circumstances: the CONSERVE 2021 statement. *JAMA*. 2021;326(3):257-265.
40. Kayambankadzanja RK, Schell CO, Gerdin Wärnberg M, et al. Towards definitions of critical illness and critical care using concept analysis. *BMJ Open*. 2022;12(9):e060972.
41. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601-609.
42. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. *Stat Med*. 2017;36(8):1203-1209.
43. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
44. Auriemma CL, Taylor SP, Harhay MO, Courtright KR, Halpern SD. Hospital-free days: a pragmatic and patient-centered outcome for trials among critically and seriously ill patients. *Am J Respir Crit Care Med*. 2021;204(8):902-909.
45. Mullens W, Dauw J, Martens P, et al. Acetazolamide in acute decompensated heart failure with volume overload. *N Engl J Med*. 2022;387(13):1185-1195.
46. Frank L, Basch E, Selby JV; for the Patient-Centered Outcomes Research Institute. The PCORI perspective on patient-centered outcomes research. *JAMA*. 2014;312(15):1513-1514.
47. Kersting C, Kneer M, Barzel A. Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity. *BMC Health Serv Res*. 2020;20(1):596.
48. Harhay MO, Wagner J, Ratcliffe SJ, et al. Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med*. 2014;189(12):1469-1478.
49. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. *Perspect Clin Res*. 2015;6(3):169-170.
50. Jones PW, Beeh KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med*. 2014;189(3):250-255.